

Cuando la IA piensa por nosotros: el riesgo de no aprender a razonar

Por Boris Sánchez Molano · 2026-09-24

INTRODUCCIÓN

En el *Fedro*, Platón pone en boca de Sócrates una objeción a la escritura: confiar el saber a signos externos debilitaría la memoria y daría a las personas “la apariencia de la sabiduría, no su realidad”. La objeción resultó, en gran medida, equivocada; la escritura amplió la cultura humana como pocas tecnologías. Pero contenía una intuición que merece recuperarse: cada vez que trasladamos una función mental a un soporte externo, cambiamos algo en quien la ejercía.

De ahí arranca este artículo. Durante siglos hemos delegado tareas cognitivas en herramientas (el ábaco, la calculadora, la agenda, el GPS) sin que ello erosionara nuestra capacidad de pensar; al contrario, esas herramientas liberaron recursos para tareas más complejas. La inteligencia artificial (IA) generativa introduce, sin embargo, una diferencia, ya no delegamos el cálculo o el almacenamiento de datos, sino operaciones que considerábamos el núcleo del trabajo intelectual: razonar, argumentar, sintetizar, redactar, decidir.

El problema cuesta verlo a primera vista, y ahí reside su dificultad. Un estudiante que entrega un ensayo redactado con IA presenta un producto correcto, a veces excelente. Pero el aprendizaje ocurre en el proceso de construirlo, en el trabajo de ordenar una idea, descartar fuentes, conectar conceptos y corregir los propios errores. Cuando ese proceso se externaliza o delega, el resultado mejora y el aprendizaje se vacía. A esa disociación la llamaremos la **paradoja del rendimiento**: rendir mejor no equivale a aprender más.

Este artículo no es una advertencia tecnófoba ni una defensa de prohibir la IA en las aulas. La IA generativa puede potenciar el aprendizaje y ya lo hace en muchos contextos. La tesis es más específica, la IA generativa no es pedagógicamente neutra, y la frontera entre una delegación que libera capacidades y una que las atrofia no la fija la tecnología, sino el diseño didáctico que la rodea. Donde no hay diseño, la herramienta tiende, por defecto, hacia la sustitución.

El destinatario es el cuerpo docente, en particular el de educación media y superior, niveles donde estos usos de la IA ya son cotidianos. El recorrido es el siguiente: la primera parte define qué es delegar y por qué el esfuerzo importa; la segunda examina qué cambia con la IA generativa; la tercera describe las consecuencias (deuda epistémica y pérdida de agencia); la cuarta plantea qué puede y debe hacer la docencia.

Parte I. Entender la delegación cognitiva

1. Qué es la delegación cognitiva y por qué no es nueva

La psicología cognitiva tiene un nombre preciso para este fenómeno: **delegación cognitiva** o *cognitive offloading*, definida como el uso de una acción física o de un recurso externo para reducir las demandas cognitivas de una tarea (Risko & Gilbert, 2016). Anotar una cita en la agenda, programar una alarma o guardar un número en el teléfono son formas de delegación; sacamos información o esfuerzo de la mente y lo colocamos en el entorno.

Visto así, delegar es una estrategia adaptativa. La memoria de trabajo (el espacio mental donde manipulamos información mientras pensamos) tiene una capacidad limitada, y descargar parte de la carga en soportes externos permite destinar esos recursos a operaciones más exigentes. La historia de la cultura puede leerse, en parte, como una historia de delegaciones sucesivas, de la tradición oral a la escritura, del cálculo manual a los buscadores.

Con la llegada de los motores de búsqueda apareció lo que se conoce como efecto Google, la tendencia a recordar peor un contenido cuando esperamos poder consultarlo después, y a recordar mejor, en cambio, dónde encontrarlo (Sparrow et al.,

2011). Internet pasó a funcionar como una memoria transactiva, un almacén externo de conocimiento que sabemos consultar sin necesidad de retenerlo. La dirección del cambio resulta reconocible para cualquier docente; dejamos de preguntarnos qué sabemos para preguntarnos dónde está.

Hasta aquí, nada es alarmante. Calculadora, GPS y buscador externalizan funciones mentales (una operación aritmética, una ruta, la recuperación de un dato) y, en general, lo hacen bien. La pregunta no es si delegamos, porque siempre lo hemos hecho, sino qué delegamos y quién conserva el control del proceso. Una calculadora no decide qué problema vale la pena resolver; un GPS no elige el destino. El usuario sigue siendo el autor de la tarea, define el objetivo, interpreta el resultado, asume sus consecuencias. Esa distinción entre delegar una función y delegar la autoría del pensamiento es la que la IA generativa pone en cuestión.

2. El cerebro ahorra\, el aprendizaje exige

Para entender por qué cierta delegación es problemática, hay que detenerse antes en cómo aprende una persona. Una advertencia previa: este artículo aborda el problema desde la perspectiva pedagógica, la del desarrollo cognitivo que la educación tiene la tarea de promover, y deja la neurociencia fuera de su alcance. Bastan, para el argumento, dos ideas bien establecidas.

La primera es que el cerebro tiende a ahorrar esfuerzo. La psicología cognitiva describe a la mente humana como inclinada por defecto a la respuesta rápida y automática, y reacia al razonamiento deliberado y costoso (Kahneman, 2011). Es una economía evolutiva sensata. Pensar a fondo consume recursos y el organismo prefiere no gastarlos sin necesidad. El problema es que el aprendizaje profundo se ubica del lado del esfuerzo costoso. Si se deja a su inclinación natural, la mente toma el atajo, y buena parte de la tarea educativa consiste en impedir ese atajo cuando el atajo cancela el aprendizaje.

La segunda idea es la de las **dificultades deseables** (Bjork, 1994; Bjork & Bjork, 2011). Hay obstáculos que, lejos de estorbar, son la condición del aprendizaje duradero. Recuperar un contenido de memoria en lugar de releerlo, espaciar la práctica en el tiempo, enfrentar problemas que aún no se dominan, todo ello hace que

aprender se sienta más difícil y lento, y por eso mismo produce un conocimiento más sólido y más transferible. De ahí una distinción que merece retenerse, el rendimiento durante la práctica y el aprendizaje son cosas distintas. Una tarea puede salir fluida y dejar poco; una tarea ardua puede salir trabada y enseñar mucho.

Releer apuntes genera una sensación de fluidez (el material resulta familiar, se reconoce sin esfuerzo) que el estudiante interpreta con facilidad como comprensión. Pero familiaridad no es dominio. La fluidez es una señal poco fiable del aprendizaje, y la mente la confunde una y otra vez con saber. Esta observación reaparecerá más adelante: una respuesta de IA, redactada con una fluidez impecable, activa el mismo malentendido.

Cuando la delegación se vuelve hábito sin que nadie lo haya decidido, el abandono del esfuerzo adquiere una forma reconocible: el **sedentarismo cognitivo** (Sigman & Bilinkis, 2023). La metáfora es más precisa que la mera atrofia, el sedentarismo físico no duele en el momento; duele después, cuando el cuerpo necesita responder a una demanda que ya no puede sostener. El sedentarismo cognitivo funciona igual. La mente no percibe la pérdida mientras la herramienta esté disponible. La percibe cuando no lo está, ante el examen, el proyecto real, el problema que el modelo no resuelve igual. La atrofia de la atención sostenida y del esfuerzo deliberado es su síntoma más cotidiano, aunque no su diagnóstico más grave.

Un último concepto ayuda a precisar cuándo delegar ayuda y cuándo perjudica. La teoría de la carga cognitiva distingue entre la **carga extrínseca** (el esfuerzo que exige todo lo accesorio a la tarea, el formato, la ortografía de las citas, lo procedimental) y la **carga intrínseca** (el esfuerzo propio de la operación que constituye el aprendizaje, estructurar un argumento, decidir qué es relevante, evaluar alternativas) (Sweller et al., 2019). Externalizar la carga extrínseca suele ser beneficioso; libera memoria de trabajo y atención que el estudiante puede reinvertir en lo importante. *Externalizar la carga intrínseca equivale a entregar el aprendizaje mismo.*

De ahí la clave pedagógica central. La **fricción cognitiva** (el esfuerzo, la duda, el error, el trabajo de dar forma a una idea) es el lugar donde el aprendizaje ocurre, aunque la tecnología tienda a tratarla como un defecto del proceso que hay que

eliminar. Cuando una herramienta retira esa fricción, además de ahorrar trabajo puede retirar la oportunidad de desarrollo que la educación existe para ofrecer. El riesgo, formulado en términos pedagógicos, no es que la mente se “dañe”, sino algo más silencioso, que la construcción cognitiva no llegue a producirse.

Parte II. El salto cualitativo de la IA generativa

3. De la herramienta al motor cognitivo

La pregunta con que cerró la primera parte (qué delegamos y quién conserva el control del proceso) permite situar la novedad de la IA generativa.

Hasta hace poco, la incorporación de tecnología digital en el aula había operado bajo lo que cabe llamar un archipiélago tecnológico: el laboratorio de informática, el procesador de texto y la plataforma virtual funcionaban como islas, sin un flujo pedagógico coherente que las conectara. El debate sobre esas herramientas quedaba atrapado en la ferretería tecnológica (la obsesión reduccionista por qué dispositivo usar, cómo redactar el prompt perfecto, cómo detectar el plagio), sin trasladar la pregunta al territorio que importa, para qué enseñamos y cómo aprende el estudiante. La IA generativa ofreció resolver ese problema de fragmentación al proponer un ecosistema integrado y con memoria contextual. El problema es que ese mismo ecosistema fluido e integrado tiene, por su diseño, una consecuencia que conviene examinar en detalle, y examinarla empieza por aclarar qué es, y qué no es, un modelo de lenguaje.

Un modelo de lenguaje de gran escala funciona de un modo distinto al de una base de datos. En lugar de almacenar hechos para recuperarlos cuando se le consulta, aprende regularidades estadísticas del lenguaje a partir de enormes cantidades de texto y, ante una entrada, genera la secuencia de palabras más probable. Bender y sus colegas (2021) propusieron para este funcionamiento la imagen del “loro estocástico”: un sistema que reproduce con notable destreza las formas del lenguaje sin un modelo del mundo, sin intención y sin comprensión de lo que enuncia. La expresión tiene aquí un sentido técnico antes que descalificatorio, el modelo produce texto plausible, y la plausibilidad no requiere comprensión.

Esa precisión cambia la naturaleza de la delegación. Un buscador devuelve una lista de enlaces, y el usuario todavía debe leer, descartar, contrastar y sintetizar; es decir, debe hacer el trabajo cognitivo. Un modelo generativo devuelve una respuesta ya sintetizada y redactada. La operación intelectual en la que residía el aprendizaje (ordenar, jerarquizar, conectar, formular) viene hecha de fábrica.

Frente al estudiante, un modelo de lenguaje opera como un **motor cognitivo**, un sistema que simula el razonamiento y entrega su producto terminado. La máquina no piensa en ningún sentido relevante, y el riesgo tampoco está en su mente, que es un asunto filosófico. Está en la división del trabajo cognitivo que la herramienta instala por defecto, su salida ocupa el lugar donde habría ido el pensamiento del estudiante.

La calculadora externalizaba una función (el cálculo) y dejaba intactas la definición del problema y la interpretación del resultado. El modelo generativo puede absorber el problema entero. *Cuando lo hace, el usuario deja de ser el autor del proceso y pasa a ser su receptor.*

4. Del offloading al outsourcing

No toda delegación en la IA es problemática; sostenerlo sería tan inexacto como negar el problema. La distinción entre carga extrínseca e intrínseca ofrece el criterio para separar una de otra.

Cuando un estudiante pide a la IA que corrija la ortografía de un texto, ajuste el formato de las citas o genere un ejemplo rutinario de un concepto que ya domina, externaliza carga extrínseca. Esa delegación es el offloading descrito por Risko y Gilbert (2016), y puede ser pedagógicamente útil, libera atención y memoria de trabajo para reinvertirlas en lo que sí importa.

Cuando, en cambio, pide a la IA que estructure la argumentación, decida qué fuentes son relevantes, formule la tesis o construya el marco conceptual del trabajo, externaliza carga intrínseca, entrega las operaciones en las que consistía el aprendizaje. Esa delegación es de otra naturaleza, la que llamaremos **outsourcing epistémico**, la subcontratación del pensamiento mismo.

La literatura reciente ya tiene un nombre para ella: **rendición cognitiva** (*cognitive surrender*), definida como la tendencia a deferir el juicio, el esfuerzo y la responsabilidad a la salida del sistema sin evaluación crítica (Shaw & Nave, 2026). La distinción respecto al *offloading* es precisa, el *offloading* es una delegación estratégica que mantiene activo el razonamiento interno; la rendición cognitiva es una abdicación del razonamiento mismo. Shaw y Nave la documentaron en tres experimentos preregistrados, los participantes adoptaron las respuestas del modelo en más del 80% de los ensayos donde la IA era incorrecta, y su precisión cayó 15 puntos porcentuales por debajo de la línea base sin IA. El outsourcing epistémico es, en términos conductuales, la rendición cognitiva en contexto de aprendizaje.

El problema está en que la frontera entre ambos tipos de delegación no está señalizada y el tránsito suele ser gradual e invisible. La interfaz conversacional y la fluidez de las respuestas generan una sensación de control que encubre algo concreto, decisiones sobre qué es relevante, qué cuenta como buena explicación, qué solución es aceptable, están siendo tomadas por el sistema. *El estudiante cree estar “usando una herramienta” cuando ya ha cedido, sin advertirlo, la autoría del proceso.*

Por eso el asunto es pedagógico y excede el buen o mal uso individual. Un estudiante rara vez decide, en un momento identificable, “a partir de aquí dejo de pensar”. El desplazamiento ocurre sin que se perciba, y lo que no se percibe no puede regularse. Si el diseño de la tarea no marca esa línea, el estudiante, solo ante la herramienta, carece de un criterio fiable para localizarla.

5. Sesgo de automatización e ilusión de profundidad

¿Por qué ese desplazamiento es tan fácil de dar y tan difícil de detectar? Dos mecanismos cognitivos bien documentados, ambos anteriores a la IA, lo explican.

El primero es el **sesgo de automatización**: la tendencia a confiar en exceso en los sistemas automatizados, a aceptar su salida y a reducir la vigilancia sobre ella (Parasuraman & Manzey, 2010). Este sesgo se estudió durante décadas, y la investigación arroja un dato que merece atención, aparece tanto en personas inexpertas como en expertas, y no se elimina con la simple práctica ni con instrucciones de cautela. La consecuencia para la docencia es directa, pedir a los

estudiantes “que sean críticos” es necesario, pero insuficiente por sí solo, porque el sesgo opera por debajo de la buena voluntad. En una encuesta a trabajadores del conocimiento, una mayor confianza en la IA se asoció con un menor ejercicio del pensamiento crítico, mientras que una mayor confianza en las propias capacidades se asoció con uno mayor (Lee et al., 2025). *La delegación desplaza el esfuerzo y, con él, la disposición a hacerlo.*

El diseño de las herramientas de IA amplifica ese sesgo. Las interfaces se construyen bajo el principio de la fluidez sin fricción; la respuesta inmediata y pulida es una virtud comercial que promueven las buenas prácticas de UX/UI. Una respuesta bien escrita y aplomada se acepta sin activar el trabajo analítico que haría falta para verificarla, la fluidez se lee como validez. La calidad de la redacción no dice nada sobre la solidez del contenido, y un modelo de lenguaje está optimizado para producir texto fluido.

El segundo mecanismo es la **ilusión de profundidad explicativa** (Rozenblit & Keil, 2002). Las personas creen comprender los fenómenos complejos con mucha más precisión de la que en realidad poseen, y la ilusión se desploma en el instante en que se les pide producir una explicación causal, paso a paso, de aquello que creían entender. En el campo pedagógico este mecanismo tenía ya un nombre más intuitivo, la **ilusión de competencia**, la falsa sensación de dominio que produce el contacto fácil con un material que se reconoce pero no se ha internalizado; con la IA, ese contacto es inmediato y la respuesta llega ya elaborada. Ambas formulaciones describen el mismo proceso con distinto énfasis, la psicología cognitiva lo ancla en el conocimiento explicativo; la pedagogía lo sitúa en el punto de uso, en el momento en que el estudiante decide, con toda convicción, que ya sabe.

La combinación de ambos mecanismos describe la trampa con claridad. La IA entrega una explicación coherente y sin contradicciones aparentes. Leerla produce la sensación de haber entendido, y el estudiante evaluaría su propia comprensión como alta. Pero la explicación fue construida por el sistema; el estudiante no edificó el modelo causal ni recorrió el camino. Tiene el texto, no la comprensión. Elsayed y Verheyen (2024) empezaron a examinar este efecto en la interacción con ChatGPT, en un terreno de estudio todavía inicial.

Hay, además, un efecto de cierre. Una respuesta perfectamente articulada satisface de forma prematura la necesidad de resolver la pregunta, apaga el impulso de seguir indagando. La duda, identificada antes como parte de la fricción donde ocurre el aprendizaje, es lo que la coherencia del texto suprime. El pulido de la salida desactiva la inquietud que habría conducido a aprender.

El resultado es una persona que siente que sabe, que rinde como si supiera, pero que no ha construido lo que saber requiere. Esa distancia entre la comprensión sentida y la comprensión real persiste, y se acumula.

Parte III. La deuda epistémica y la pérdida de agencia

6. Atrofia de habilidades

El efecto más visible, y a la vez más limitado, es también el más fácil de nombrar. La literatura sobre tecnología y cognición describe desde hace tiempo el **deskilling** (la atrofia o descualificación de habilidades); *cuando una tarea cognitiva se externaliza de forma sostenida, la capacidad de realizarla sin ayuda se debilita*. Es el fenómeno que se invoca al señalar que el uso constante de la calculadora puede debilitar la fluidez del cálculo mental, o el del GPS, el sentido de la orientación. El sedentarismo cognitivo descrito en la primera parte es la cara cotidiana de este proceso, la atrofia lenta de la disposición a ejercer una destreza, más que su pérdida repentina.

El deskilling es un riesgo real, pero merece dos matices.

El primero es que el proceso no es determinista. Las mismas herramientas que descualifican en un contexto pueden, en otro, liberar al usuario de tareas rutinarias y permitirle concentrarse en operaciones de orden superior, con un efecto de signo contrario. Que una tecnología erosione o potencie una habilidad depende de cómo se inserte en la tarea. Afirmar sin más que la IA “nos hará menos capaces” es tan impreciso como la afirmación contraria.

El segundo matiz, más importante para la educación: el deskilling describe la pérdida de una habilidad que ya se tenía o se estaba desarrollando. Pero el estudiante, y sobre todo el estudiante novato, no está perdiendo una destreza consolidada; está en riesgo

de no construirla nunca. Más allá de que una capacidad existente se atrofie, el problema pedagógico de fondo es que la comprensión que una tarea debía construir no llegue a edificarse.

Para nombrar ese saber no construido y la erosión, más profunda, de la capacidad de ser autor del propio conocimiento, hacen falta dos conceptos más precisos que el deskilling.

7. Deuda epistémica

La **deuda epistémica** es el saber no construido que se acumula cuando se delega de forma sistemática el pensamiento en la IA. El estudiante obtiene un beneficio inmediato, un producto correcto entregado a tiempo, pero la comprensión que ese producto debía exigir no se edificó. Esa carencia no se manifiesta de entrada. Se manifiesta después, cuando aparece un problema nuevo que el modelo no resuelve igual, una variación que exige adaptar lo aprendido, un error que hay que diagnosticar, o una situación (un examen, una intervención profesional) en la que la herramienta no está disponible.

Siempre hemos dependido del conocimiento de otros. Cuando un estudiante se apoya en un experto humano (un profesor, un texto, un especialista), existe una cadena de justificación que puede recorrerse, el experto explica el porqué, responde de lo que afirma, puede ser interrogado. Cuando se apoya en un modelo de lenguaje, recibe una salida producida sin comprensión y sin una secuencia causal trazable, el sistema no “sabe” por qué, no responde de nada y puede equivocarse con la misma fluidez con que acierta. El estudiante tiene el resultado (funciona, es entregable), pero no lo posee cognitivamente.

De ahí surge el **experto frágil**, quien produce textos solventes y soluciones que funcionan, pero carece de la base para intervenir cuando algo falla o cambia. *Su competencia es real mientras el contexto sea estable y la herramienta esté disponible.* La fragilidad solo se revela ante la anomalía. Es la cristalización del sedentarismo cognitivo sostenido, alta superficie, base hueca.

Con esto la distinción respecto al deskilling queda nítida, la atrofia describe una habilidad que se tuvo y se perdió; la deuda epistémica, una comprensión que nunca se construyó.

8. La agencia epistémica

La deuda epistémica es una consecuencia. El concepto que subyace es la agencia epistémica.

La **agencia epistémica** es la capacidad de una persona de iniciar, regular y evaluar críticamente sus propios procesos de formación de creencias y de construcción de conocimiento. Consiste en ejercer un juicio propio sobre la información que llega de fuentes externas (recibirla lo hace cualquiera), en decidir qué cuenta como evidencia, qué afirmaciones merecen crédito, cuándo una creencia debe revisarse. Es, en otras palabras, una forma de autonomía intelectual.

Esa autonomía se apoya en una capacidad concreta y bien estudiada, la metacognición (Flavell, 1979), el monitoreo y la regulación que una persona ejerce sobre su propio conocer: saber qué sabe, qué no, y qué hace falta para saberlo.

El riesgo más profundo de la delegación no estructurada no es la pérdida de una habilidad ni la deuda en sí, sino algo más difícil de recuperar, la erosión de la disposición y la capacidad de ser autor del propio conocimiento. La erosión opera por tres vías.

La primera es el descenso del *estándar de justificación*. El sesgo de automatización hace que el estudiante acepte la salida del modelo sin interrogarla; la pregunta elemental (“¿por qué esta respuesta y no otra?”) se omite. El criterio de “esto es correcto” se sustituye, sin que medie decisión consciente, por “esto lo dijo el sistema”.

La segunda es el desplazamiento de la *autoridad epistémica*. El punto de referencia deja de ser el juicio del propio sujeto (“¿me convence? ¿lo he comprendido?”) y pasa a ser el resultado del modelo. La autoridad sobre qué es verdadero migra del estudiante al sistema, y lo hace de forma silenciosa, porque la interfaz conserva la apariencia de que el usuario sigue al mando.

La tercera es más sutil, un *currículo oculto*. Cada interacción con la IA enseña, de forma implícita, una idea de lo que es pensar bien. Y lo que premia es la certeza sobre la ambigüedad y la rapidez sobre la deliberación: la respuesta llega de inmediato, pulida y sin fisuras. El estudiante aprende, sin que nadie se lo enseñe de manera explícita, que el buen pensamiento es veloz y concluyente, justo lo contrario de lo que la indagación genuina, lenta y tentativa, realmente es.

Hay un matiz necesario. Bien diseñada, la IA puede fortalecer la agencia epistémica en lugar de erosionarla; puede funcionar como un interlocutor que confronta, que obliga a justificar, que introduce la duda en lugar de cancelarla. Todo depende de quién conserve el control reflexivo del proceso. Pero hay que ser claros sobre el punto de partida, librada a su diseño por defecto, fluido y complaciente, la IA generativa empuja hacia la erosión. Revertir esa inclinación exige un trabajo deliberado, y ese trabajo ocupa la cuarta parte.

9. Por qué cuesta sostener la agencia

Podría pensarse que basta con la voluntad, un estudiante que decida seguir siendo el autor de su pensamiento lo conseguirá. La realidad es más difícil.

Sostener la agencia epistémica cuesta porque el entorno entero trabaja en su contra. El problema excede a la IA. El conjunto del ecosistema digital está diseñado en torno a la fluidez y la inmediatez: la interfaz que se anticipa a la intención, la recomendación que evita la búsqueda, la notificación que dirige la atención. La fricción se trata, en todas partes, como un defecto de diseño que hay que eliminar.

El cerebro ya prefiere, por economía, el atajo. Lo nuevo es que ahora el entorno retira de forma sistemática los obstáculos que antes hacían ese atajo menos accesible. Sostener la agencia significa, en estas condiciones, elegir a contracorriente el camino más costoso, tarea a tarea.

La preocupación que sigue es una proyección razonable, no un hecho establecido. Si el esfuerzo intelectual sostenido exige resistir un entorno optimizado para que no lo hagamos, ese esfuerzo corre el riesgo de volverse un bien escaso, ejercido solo por quienes dispongan del tiempo, la formación y la disposición para resistir. El pensar con esfuerzo podría dejar de ser una práctica común para convertirse en una suerte de

bien de lujo.

La agencia no puede sostenerse apelando a la voluntad individual del estudiante.

Pedirle que “se esfuerce más” o que “use bien la herramienta” deposita sobre él una carga que el diseño del entorno vuelve casi imposible de levantar en solitario. *La agencia epistémica tiene que ser construida y protegida de forma deliberada. Diseñar los entornos donde eso ocurra es, en este punto, la tarea del docente.*

10. Soberanía intelectual

Si la agencia epistémica es la capacidad, la **soberanía intelectual** es la postura que la mantiene activa: seguir siendo quien fija las metas del aprendizaje, quien decide qué cuenta como evidencia, quien determina cuándo una creencia se acepta o se revisa. Los algoritmos pueden ofrecer datos, patrones y formulaciones; pero la dirección del saber, qué problemas importan y para qué, debe seguir siendo humana.

La contrafigura del experto frágil es el **prosumidor cualificado**, quien consume información algorítmica para deconstruirla, verificarla y resignificarla desde su propio contexto. El prosumidor cualificado toma la salida de la IA como punto de partida y la somete a interrogación, como haría con cualquier interlocutor. Esa actitud se construye mediante práctica, modelado y diseño de tareas que la exijan; la buena intención no alcanza.

Frente a la delegación no estructurada, la soberanía intelectual descansa en dos disposiciones concretas, ambas exigentes.

La primera es la *vigilancia metacognitiva*: el hábito de monitorear el propio estado de comprensión y, en particular, de distinguir dos cosas que la fluidez confunde. Una es “tengo delante un texto correcto”; otra, muy distinta, es “he comprendido esto”. La vigilancia metacognitiva consiste en preguntarse, de forma deliberada, antes de dar una tarea por terminada, ¿podría explicar esto sin la herramienta?, ¿podría reconstruir la cadena causal, paso a paso? Es la prueba que desactiva la ilusión de competencia.

La segunda es el *escepticismo crítico*: tratar la salida de la IA como una afirmación que hay que examinar antes de aceptarla. Verificar, contrastar con otras fuentes, buscar lo que falta o lo que está mal, atender a aquello que el modelo presenta con

seguridad pero no puede justificar. La desconfianza generalizada sería tan inútil como la credulidad; lo que se pide es un escrutinio calibrado.

No hay que idealizar estas estrategias. Son costosas, exigen el gasto de esfuerzo que la mente y el entorno nos empujan a evitar, y van a contracorriente. Por eso no surgen de manera espontánea. La vigilancia y el escepticismo se aprenden y se sostienen mediante una estructura que los exija. Ese costo tiene una contrapartida; la soberanía intelectual es lo que separa al sujeto de la manipulación. Quien no examina lo que acepta queda subordinado a quien lo produjo.

Sostener la soberanía intelectual y dominar la herramienta son cosas distintas. Enseñar a redactar mejores prompts es una habilidad operativa que, por sí sola, puede profundizar la dependencia, vuelve al usuario más eficaz extrayendo respuestas sin volverlo más capaz de juzgarlas. Comprender qué es y cómo funciona un modelo de lenguaje (sus límites, su falta de comprensión real, su tendencia a la fluidez) protege más la soberanía intelectual que cualquier catálogo de instrucciones eficaces.

Lo que está en juego en la delegación cognitiva no es la eficiencia, sino la autoría. Que los estudiantes usarán la IA está fuera de duda; la cuestión es si seguirán siendo los autores de su propia comprensión mientras la usan. Como esa autoría no se sostiene sola en un entorno diseñado en su contra, la pregunta se traslada al diseño de la enseñanza.

Parte IV. El docente como arquitecto

11. El docente como arquitecto del aprendizaje

Trasladada la pregunta, el foco pasa del estudiante al diseño de la enseñanza, que debe ofrecer a la agencia epistémica un lugar donde construirse. Con ello cambia la figura del docente.

Durante mucho tiempo, el rol del docente se entendió como transmisor de contenidos y, en el mejor de los casos, orquestador de productos: explicar, asignar, recibir, calificar. Esa función ya estaba siendo cuestionada antes de la llegada de la IA generativa (los enfoques constructivistas, el aprendizaje basado en proyectos y la

evaluación formativa ya habían empezado a desplazarla). La IA vuelve inviable seguir aplazando ese desplazamiento. Si los productos que antes evidenciaban aprendizaje ahora pueden generarse sin él, el docente ya no puede limitarse a recibirlos, tiene que diseñar las condiciones bajo las cuales el aprendizaje ocurra.

La nueva tarea pide una imagen distinta, la del **docente arquitecto**. Un arquitecto no entrega un objeto terminado; diseña el espacio en el que algo va a suceder. Aplicado a la enseñanza, el arquitecto del aprendizaje diseña cuatro cosas: (1) *sentido*, la razón por la cual la tarea importa más allá de cumplirla; (2) *reto*, la exigencia cognitiva que no permite resolverse por simple delegación; (3) *experiencia*, la secuencia vivida (prueba, error, ajuste, decisión) en la que ocurre el aprendizaje; y (4) *fricción productiva*, las dificultades deseables (Bjork & Bjork, 2011) que el diseño introduce de forma deliberada para que la construcción cognitiva tenga lugar.

El objetivo último de ese diseño es formar **arquitectos de sentido**, personas capaces de conectar saberes fragmentados, dotar de valor ético a la técnica y dar dirección humana a la precisión algorítmica. El docente arquitecto diseña el espacio; el estudiante, al recorrerlo con esfuerzo genuino, desarrolla la capacidad de construir su propio saber. Esa es la cadena que la delegación no estructurada interrumpe.

La pregunta no es nueva. Markauskaite y sus colegas (2022) la formularon en términos explícitos (¿qué capacidades necesita un aprendiz en un mundo saturado de IA?) y sostuvieron que responderla exige repensar la relación entre aprendizaje humano y máquinas inteligentes. Lo que se propone aquí se inscribe en esa línea y se concentra en los ejes que estructuran la nueva tarea del docente.

Este giro hace más exigente la tarea docente. Le pide conocer su disciplina (eso no cambia) y además saber cómo se aprende, cómo funcionan las herramientas que sus estudiantes ya usan, qué tipo de tareas resisten la delegación no estructurada y cuáles la invitan. Y le pide aceptar que su trabajo más importante está ahora en el diseño de lo que ocurre entre la clase y la corrección. Las secciones siguientes desarrollan esa tarea, empezando por lo más urgente: el peso del resultado.

12. Desplazar el peso del resultado

La pedagogía de las últimas décadas se ha orientado, de forma creciente, al producto final. El ensayo, el informe, el mapa conceptual, la presentación, la entrega, lo que el estudiante “saca” es lo que se ve, lo que se corrige y, en buena medida, lo que se califica. Este sesgo hacia el resultado no nació con la IA; lo precede. La era de Google ya había desplazado las tareas escolares hacia la síntesis (resumir, comparar, integrar), pero las evaluó casi siempre como producto; importaba cómo quedaba la síntesis y poco cómo se había hecho.

Ese diseño era ya imperfecto, porque ofrecía a la fluidez digital una entrada cómoda al aula. Con la IA generativa, sin embargo, se vuelve insostenible. Un producto bien estructurado, gramaticalmente impecable y con referencias plausibles puede generarse sin que haya mediado prácticamente ninguna construcción cognitiva por parte de quien lo entrega. *Si el peso sigue sobre el producto, el resultado es predecible, pensar se vuelve opcional. Y lo opcional, en un entorno sin fricción, tiende a no hacerse.*

Cuando las calculadoras se generalizaron en las aulas, los docentes de matemáticas vivieron una crisis similar a la actual. La herramienta entregaba el resultado correcto con un mínimo esfuerzo, e incluso las soluciones aparecían disponibles en pantalla. Muchos exámenes, cuyo valor descansaba en obtener ese resultado, quedaron obsoletos. El debate de aquellos años se planteó, al comienzo, en términos de prohibir o permitir. La respuesta que terminó imponiéndose fue otra y mucho más fértil, desplazar el peso del resultado al procedimiento.

Las pruebas pidieron, cada vez más, mostrar la hoja de procedimientos, la secuencia de pasos, las decisiones intermedias, la justificación de por qué cada operación seguía a la anterior. El número final dejó de ser, por sí mismo, prueba de nada. Lo que evidenciaba la comprensión era el camino visible.

En el contexto de la IA generativa, esa lógica tiene una versión más amplia y pedagógicamente más precisa, la **trazabilidad del pensamiento**. Documentar el proceso de construcción de ideas (borradores, iteraciones, diálogos con el modelo, decisiones tomadas y descartadas, reflexión metacognitiva sobre lo entendido y lo no entendido) tiene poco de requisito burocrático: es la evidencia más directa de que la construcción cognitiva ocurrió. Cuando solo hay un producto, la calidad del texto

entregado no prueba, por sí misma, que haya habido aprendizaje.

La analogía con la calculadora no es perfecta (los productos textuales son más difíciles de descomponer en pasos discretos que un cálculo aritmético), pero la lección pedagógica se traslada. La conclusión es técnica, sin nostalgia ni elogio de la lentitud por sí misma. Si el producto ha dejado de ser un indicador fiable del aprendizaje, el diseño debe producir otros indicadores. Reducir el peso del resultado no es bajar la exigencia; es ponerla donde el aprendizaje efectivamente sucede.

13. Rediseñar tareas y evaluación

El desplazamiento se concreta en tres cambios: las premisas con que se diseña la tarea, las tareas mismas y la evaluación.

El cambio en las premisas. La primera decisión arquitectónica es asumir la presencia de la IA. No prohibirla (la prohibición es inútil y abandona al estudiante a una integración no guiada en cuanto sale del aula), pero tampoco ignorarla. Una tarea bien diseñada en este entorno es una que no se resuelve de manera significativa por delegación pasiva; una que exige el juicio situado del estudiante, su experiencia, su conocimiento del contexto, sus decisiones documentadas. El criterio operativo es claro, si la tarea puede completarse a un nivel aceptable copiando la primera salida de un modelo, la tarea está mal diseñada para el momento que vivimos. Es una premisa de diseño más que una regla absoluta (hay aprendizajes en los que esa primera salida puede ser un punto de partida legítimo).

El cambio en la tarea misma. Sobre esa premisa, las tareas incorporan fricción productiva. En términos concretos, esto significa diseñar tareas que pidan al estudiante construir un argumento desde una posición que la respuesta por defecto del modelo no asume; identificar qué falta o qué está mal en una salida de IA; documentar la iteración (qué se intentó, qué se descartó, por qué); defender oralmente lo que se entregó por escrito; tomar y justificar decisiones en un contexto particular que el modelo no conoce. La fricción no consiste en hacer la tarea más difícil en términos absolutos, sino en hacer que el pensamiento del estudiante sea aquello que la tarea efectivamente requiere.

El cambio en la evaluación. Las dos transformaciones anteriores pierden su fuerza si la evaluación sigue mirando solo al producto. Evaluar el proceso tiene una larga tradición (la evaluación formativa, interesada por el aprendizaje mientras ocurre, lleva décadas argumentada y documentada [Black & Wiliam, 1998]), y la IA generativa convierte su adopción en una necesidad. Evaluar el proceso significa atender a la trazabilidad del pensamiento, los borradores, las preguntas, las justificaciones, el diálogo con la herramienta. Significa también incorporar formatos que el producto entregado no puede sustituir: defensas orales, explicaciones sin ayuda, aplicación del aprendizaje a una variación nueva, evaluación entre pares. La consigna es molesta pero clara, un docente que solo califica el texto final ha delegado, sin saberlo, su juicio evaluativo en lo que el modelo produjo.

Estos tres cambios dependen entre sí. Una evaluación de proceso sin tareas que generen proceso no tiene nada que evaluar; tareas con fricción sin una evaluación que la reconozca pierden su sentido y se vuelven, para el estudiante, un castigo gratuito. El rediseño es sistémico, no cosmético.

14. Ambientes de aprendizaje con sentido y contexto

Las secciones anteriores presuponen un eje sin el cual lo demás no se sostiene, el sentido de la tarea.

Una observación pedagógica anterior a la IA gana aquí peso: los estudiantes delegan con más facilidad las tareas que perciben como arbitrarias o genéricas. Si la actividad no significa nada para quien la realiza (un ejercicio en el vacío, intercambiable con cualquier otro), el estudiante no pierde nada personal al externalizarla. No hay nada propio en juego, y pedirle vigilancia metacognitiva en ese contexto es pedirle un esfuerzo sin razón.

Al contrario, cuando la tarea está anclada en un contexto real, la salida genérica del modelo se vuelve visiblemente insuficiente. La IA puede producir un análisis competente del cambio climático en abstracto, pero no puede aportar la lectura situada del barrio del estudiante, la entrevista con una persona local, la observación del aula, el diagnóstico del caso particular que se está estudiando.

Conviene distinguir el sentido de la motivación entendida como decoración. No se trata de “hacer interesante” la clase; se trata de que la tarea pida del estudiante lo que no podría tener sin él. Sentido y contexto son, en ese punto, defensas de fondo contra la delegación no estructurada.

Diseñar tareas con sentido no se puede automatizar ni copiar de una guía; pide conocer al grupo concreto, el contexto local, los puntos en los que cada disciplina conecta con algo real. Es la pieza sobre la que se apoyan las otras. Sin sentido, la fricción se vive como obstáculo arbitrario; con sentido, como condición del trabajo.

15. Hacia un uso adecuado de la IA

Las decisiones de diseño anteriores permiten formular el uso de la IA en términos pedagógicos, como un conjunto reducido de principios que el diseño puede convertir en el camino de menor resistencia para el estudiante.

La IA después, no antes. El primer principio es que el esfuerzo propio del estudiante debe ocurrir antes de que la herramienta aparezca. Esbozar el argumento, intentar la solución, fijar la posición (aunque sea imperfecta) sin ayuda, y solo después usar la IA para contrastar y refinar. La secuencia importa, cuando la IA llega primero, ocupa el lugar del pensamiento; cuando llega después, lo amplía. Este principio se sostiene con diseño, con tareas cuyas primeras fases obliguen a producir algo propio que después se confronta con la herramienta.

La IA como interlocutor, no como oráculo. El segundo principio reorienta la función del modelo. Se trata de usarlo como contraargumento del propio razonamiento, pedirle objeciones, identificar lo que falta, justificar las propias decisiones. Bien usada, la IA abre la discusión en lugar de clausurarla. Aquí también el diseño es lo que cuenta, tareas que piden al estudiante registrar las objeciones del modelo y responderlas convierten al sistema en lo que mejor puede ser para el aprendizaje, un interlocutor exigente.

La IA visible, no oculta. El tercer principio es la transparencia. Lo oculto no puede evaluarse ni corregirse. Hacer visible el uso (documentar el diálogo con el modelo, marcar qué fue del estudiante y qué de la herramienta, explicar por qué se aceptó una salida o se rechazó otra) no es vigilancia: es la condición elemental para que la

trazabilidad del pensamiento pueda ser observada. Una IA invisible deja al docente sin nada que mirar y al estudiante sin nada que reconocer como propio.

Estos tres principios comparten una propiedad importante, descansan en un diseño que los vuelva el modo natural de hacer la tarea, antes que en la voluntad del estudiante. Ahí está la diferencia entre exhortar y construir agencia.

Una analogía sirve para cerrar. La IA es, para el aprendizaje, lo que un entrenador es para el atleta, puede orientar, corregir, exigir, acelerar. Pero no puede levantar el peso por él. Si lo intenta, si “ayuda” al punto de hacer el trabajo, el desarrollo que el entrenamiento busca no ocurre. La tarea del docente, en este punto, es asegurarse de que el atleta siga siendo quien levanta el peso.

CONCLUSIÓN

La delegación cognitiva no es nueva, y no toda delegación es problemática. Externalizar la carga accesoria (el formato, la corrección ortográfica, el dato puntual) libera recursos para el trabajo intelectual que sí importa, y la educación lo ha incorporado siempre que ha tenido sentido hacerlo. Lo que distingue a la IA generativa de las tecnologías anteriores no es que delegue más, sino que delega otra cosa, las operaciones (estructurar, jerarquizar, formular, justificar) en las que el aprendizaje propiamente consistía. Y lo hace por defecto, sin marcar el límite entre lo que libera y lo que sustituye.

De esa diferencia se desprenden los problemas descritos. La fluidez del producto encubre la ausencia del proceso; la coherencia de la salida desactiva la duda que conducía a comprender. El sedentarismo cognitivo se instala sin que nadie lo decida, la mente deja de ejercitarse porque el entorno retira los obstáculos antes de que puedan enseñar algo. Lo que se acumula es una comprensión nunca construida, una deuda epistémica que se paga después, ante el primer problema que la herramienta no resuelve igual. Y bajo la deuda hay algo más difícil de recuperar, la pérdida silenciosa de la agencia epistémica, la disposición y la capacidad de ser autor del propio conocimiento.

El problema no se resuelve con prohibiciones ni con catálogos de buenas prácticas. Se resuelve, si se resuelve, con diseño. Diseñar la enseñanza para que la agencia epistémica tenga dónde construirse es la nueva forma del oficio docente, más exigente que la anterior. El docente arquitecto construye entornos donde el sentido orienta, el reto exige, la experiencia ocurre y la fricción productiva hace posible el aprendizaje. Desplaza el peso del resultado al proceso, exige la trazabilidad del pensamiento, asume que la IA está presente y rediseña las tareas para que el pensamiento del estudiante sea aquello que estas efectivamente requieren. A la IA la coloca como interlocutor exigente, después del esfuerzo propio y a plena vista.

La meta es formar arquitectos de sentido: personas capaces de conectar saberes fragmentados, de dotar de valor ético y propósito humano a la precisión algorítmica, de seguir siendo los autores de su propio conocimiento en un entorno diseñado para que no lo sean. Esa formación no se produce en el producto que se entrega. Se produce en el camino que se recorre.

Cada salto tecnológico relevante (la escritura, la imprenta, la calculadora, internet) exigió a la educación rediseñarse, y cada vez la respuesta fue del mismo tipo, *más pedagogía, no más tecnología*. El rediseño que la IA generativa pide es más profundo porque lo que está en cuestión es más cercano al núcleo del pensar, pero pertenece a la misma tradición. Lo que hoy llamamos diseño educativo es, cada vez más, diseño epistémico, lo que protegemos es una manera de ser sujetos de conocimiento.

Sócrates temió que la escritura debilitara la memoria y diera a los hombres la apariencia de la sabiduría. La escritura no destruyó a la humanidad pensante; la transformó, y la educación encontró la manera de seguir formándola. La IA generativa no destruirá necesariamente la nuestra. Pero esa continuidad no está garantizada. Depende de que la docencia acepte, con la seriedad que el momento exige, que su tarea es ahora diseñar las condiciones en las que el saber pueda seguir construyéndose. Esa es la elección, y excede lo técnico. Es una elección sobre qué clase de mente queremos que la escuela siga produciendo.

* Este artículo desarrolla y aplica conceptos expuestos previamente por el autor en *Hacia una pedagogía del sentido* (Sánchez, 2024).

REFERENCIAS

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? En *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 610–623). Association for Computing Machinery. <https://doi.org/10.1145/3442188.3445922>

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. En J. Metcalfe & A. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185–205). MIT Press. <https://augmentingcognition.com/assets/Bjork1994.pdf>

Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. En M. A. Gernsbacher, R. W. Pew, L. M. Hough & J. R. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society* (pp. 56–64). Worth Publishers. https://bjorklab.psych.ucla.edu/wp-content/uploads/sites/13/2016/04/EBjork_RBjork_2011.pdf

Black, P., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 80(2), 139–148. <https://doi.org/10.1177/003172171009200119>

Elsayed, Y., & Verheyen, S. (2024). ChatGPT and the illusion of explanatory depth. En *Proceedings of the 46th Annual Conference of the Cognitive Science Society*. <https://www.proceedings.com/content/077/077494webtoc.pdf>

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist*, 34(10), 906–911. <https://doi.org/10.1037/0003-066X.34.10.906>

Gerlich, M. (2025). AI tools in society: Impacts on cognitive offloading and the future of critical thinking. *Societies*, 15(1), Article 6. <https://doi.org/10.3390/soc15010006>

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.

Kosmyna, N., Hauptmann, E., Yuan, Y. T., Situ, J., Liao, X. H., Beresnitzky, A. V., Braunstein, I., & Maes, P. (2025). *Your brain on ChatGPT: Accumulation of cognitive*

debt when using an AI assistant for essay writing task [Preprint]. arXiv.

<https://doi.org/10.48550/arXiv.2506.08872>

Lee, H.-P., Sarkar, A., Tankelevitch, L., Drosos, I., Rintel, S., Banks, R., & Wilson, N. (2025). The impact of generative AI on critical thinking: Self-reported reductions in cognitive effort and confidence effects from a survey of knowledge workers. En *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery. <https://doi.org/10.1145/3706598.3713778>

Markauskaite, L., Marrone, R., Poquet, O., Knight, S., Martinez-Maldonado, R., Howard, S., Tondeur, J., De Laat, M., Buckingham Shum, S., Gašević, D., & Siemens, G. (2022). Rethinking the entwinement between artificial intelligence and human learning: What capabilities do learners need for a world with AI? *Computers and Education: Artificial Intelligence*, 3, 100056. <https://doi.org/10.1016/j.caeai.2022.100056>

Parasuraman, R., & Manzey, D. H. (2010). Complacency and bias in human use of automation: An attentional integration. *Human Factors*, 52(3), 381–410. <https://doi.org/10.1177/0018720810376055>

Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. *Trends in Cognitive Sciences*, 20(9), 676–688. <https://doi.org/10.1016/j.tics.2016.07.002>

Rozenblit, L., & Keil, F. (2002). The misunderstood limits of folk science: An illusion of explanatory depth. *Cognitive Science*, 26(5), 521–562. https://doi.org/10.1207/s15516709cog2605_1

Sánchez, B. (2024). *Hacia una pedagogía del sentido; El docente como arquitecto*. Universidad Icesi. <https://eduteka.icesi.edu.co/articulos/hacia-una-pedagogia-del-sentido>

Shaw, S. D., & Nave, G. (2026). *Thinking - fast, slow, and artificial: How AI is reshaping human reasoning and the rise of cognitive surrender* [Preprint]. PsyArXiv. https://osf.io/preprints/psyarxiv/yk25n_v1

Sigman, M., & Bilinkis, S. (2023). *Artificial: La nueva inteligencia y el contorno de lo humano*. Debate. https://escuelasconectadaspilar.unipe.edu.ar/pluginfile.php/105597/mod_resource/content/1/SIGMAN.pdf

[%20Artificial.pdf](#)

Sparrow, B., Liu, J., & Wegner, D. M. (2011). Google effects on memory: Cognitive consequences of having information at our fingertips. *Science*, 333(6043), 776–778.
<https://doi.org/10.1126/science.1207745>

Sweller, J., van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review*, 31(2), 261–292.
<https://doi.org/10.1007/s10648-019-09465-5>